

Automatic Seismic Event Recognition and Later Phase Identification for Broadband Seismograms

by Cheng Tong* and Brian L. N. Kennett

Abstract Knowledge of the patterns of frequently observed seismic phases associated with specific distances and depths have been well developed and applied by seismologists (see, e.g., Richter, 1958; Kulhánek, 1990). However, up till now, the expertise of recognizing seismic event patterns for teleseisms has not been translated into automatic processing procedure. A new approach is developed to automate this kind of heuristic human expertise in order to provide a means of improving preliminary event locations from a single site.

An automatic interpretation system exploiting three-component broadband seismograms is used to recognize the pattern of seismic arrivals associated with the presence of a seismic event in real time accompanied by an identification of the individual phases. For a single station, such a real-time analysis can be used to provide a preliminary estimation of the location of the event. The inputs to the interpretation process are a set of features for detected phases produced by another real-time phase analyzer. The combinations of these features are investigated using a novel approach to the construction of an expert system. The automatic system exploits expert information to test likely assumptions about phase character and hence epicentral distance and depth. Some hypotheses about the nature of the event will be rejected as implausible, and for the remainder, an assessment is given of the likelihood of the interpretation based on the fit to the character of all available information.

This event-recognition procedure provides an effective and feasible means of interpreting events at all distances, and characterizing information between hundreds of different possible classes of patterns even when the observation is incomplete. The procedure is based on “assumption trees” and provides a useful tool for classification problems in which a number of factors have to be identified. The control set of expert knowledge used in testing hypotheses is maintained separately from the computational algorithm used in the assumption search; in consequence, the information base can be readily updated.

Introduction

The need for improved capabilities to detect, locate, and identify underground nuclear explosions has long been a major motivation for development of new technology in seismology. Seismic arrays were developed for this problem, and much work has been performed over the years to improve their effectiveness (see, e.g., Ringdal and Husebye, 1982). As noted by Bache *et al.* (1993), various automatic association programs (e.g., Engdahl and Gunst, 1966; Slunga, 1980) have been producing teleseismic bulletins automatically for many years, and the RONAPP program (Myk-

keltveit and Bungum, 1984) produces a bulletin of regional events from a single NORESS-type array. More recently, a knowledge-based system described by Bache *et al.* (1993), the Intelligent Monitoring System (IMS), provides the capability for automatic event location by a network of seismic stations.

These and other earlier methods have the same instrument requirement of a multiple-station network. However, for a sparse global network, as is currently planned for the monitoring of a comprehensive test ban treaty, the quality of the initial location estimate is critical to the quality of event characterization. In such a context, it is very important to be able to determine whether the arrivals detected at dif-

*Present address: Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, California 91125

ferent stations are likely to have arisen from a single event or from two or more geographically distinct events in a limited time period. Consequently, it is very helpful to be able to exploit good location estimates from single-station analysis.

Saari (1991) has described an algorithm for automatic event location by using single-site, three-component seismograms; however, that algorithm is limited to local events (less than 100 km). Our aim is to develop methods that can be used effectively from far-regional to teleseismic distances by exploiting the full information content of broadband seismograms. Such an improvement in single-station analysis should enable further improvement of the capability of the existing network analyses (such as in the IMS).

Most association programs use only a limited number of identifying phases, that is, P and S (including Pn , Pg , Sn , and Lg), while other frequently observed phases, including PcP , ScS , PKP , SKS , PP , PS , $PKiKP$, SS , etc., have not been utilized. However, knowledge of the patterns of frequently observed seismic phases associated with specific distances and depths have long been learned and applied by seismologists (see, e.g., Richter, 1958; Kulhánek, 1990). In this article, we describe the implementation of an automated procedure that exploits heuristic human expertise on the character of seismic phases. We use a novel expert system technique, based on an extension of techniques developed in artificial intelligence, which can cope with incomplete observed data. By using summary information on phase detections, the problem of event location can be solved interactively with the problem of phase identification for the later phases.

Recognition of a Seismic Event

Phase Detection and Characterization

The appearance of a seismic event on a seismogram consists of a sequence of individual seismic phases rising above a complex background of noise and minor seismic arrivals. To recognize or interpret a seismic event on a seismogram, the first step is to recognize individual phases.

Various techniques for automatic phase detection and characterization have been developed and commonly used over the years. Many phase detectors use algorithms based on the ratios of short-term to long-term averages along the seismic trace (see, e.g., Allen, 1978; Earle and Shearer, 1994). A number of studies have also endeavored to characterize the detected phases. Cichowicz (1993) described an automatic S -phase picker, which provides the capability of distinguishing S -type arrivals from P -type arrivals. Tong (1995) described how the waveform segment associated with a phase detection can be specified in terms of a set of parameters based on a model of the seismic wavelet. Tong and Kennett (1995) show how this technique can be applied and developed to a three-component analysis; they demonstrated that in the teleseismic regime, the following six at-

tributes for each detected phase can be extracted in real time; for example, for the i th phase, we would have

1. t_i , the arrival time;
2. a_i , the amplitude;
3. v_i , the local frequency;
4. ϕ_i , the azimuth in the horizontal plane;
5. ψ_i , the angle of incidence to the vertical; and
6. c_i , the P or S type of the arrival.

In the subsequent development, we assume that phase detection and characterization has been undertaken by some suitable algorithm, and the six attributes above are available for each detection. In our specific implementation, the procedures described by Tong and Kennett (1995) are employed.

We show how the sequence of detected phases can be associated with seismic events by analyzing the attributes of the phases and how the phases themselves can be identified at the same time.

The Patterns of Seismic Events

The pattern classes that we try to recognize are seismic events, and since different epicentral distance or depth can produce totally different sets of observed phases, we will endeavor to classify the events by range and depth. To provide a reasonably comprehensive coverage of possible wave fields, we work with 90 possible distances (dividing the range from 0° to 180° into 2° intervals) and a selection of four depths (0, 100, 300, and 600 km) as in the summary tables for the IASP91 travel-time tables (Kennett, 1991). In this way, with 360 pattern classes, we are able to provide a characterization into shallow, intermediate, and deep events and to provide a reasonable sampling of the dependence on epicentral distance.

If there are N_p phases in a pattern, we have a set of $6N_p$ features (with six attributes per phase) describing the pattern for an event. The problem of event recognition can then be viewed as a classification problem in a $6N_p$ -dimensional vector space, with 360 possible destination classes.

One obvious approach to the event classification problem would be to set up 360 model patterns and compare any set of observed phase attributes with all these model patterns to find the best match. However, using this approach, we will face severe difficulties. We need to recognize that the phase summaries extracted from the observed data may be distorted due the presence of noise or incomplete (e.g., the amplitude for some phase may lie below the detection threshold).

The phases associated with a particular source depth and distance will have different degrees of importance; thus, some phases must be observed for this combination to be identified, while others might or might not be observed. The attributes for each phase must also be regarded as having differences in significance, since, for example, the relative amplitudes of phases will vary depending on the source

Table 1
Models and Related Ranges

P-S Identity	Time Interval Range (min)	Distance Range (°)
<i>P-S</i>	2.18–10.43	12–85
<i>P-SKS</i>	9.37–10.65	82–99
<i>P_{diff}-SKS</i>	9.39–10.69	100–129
<i>PP-SKS</i>	3.96–7.22	82–129
<i>PP-PS</i>	9.20–10.14	104–125
<i>PKP-SKS</i>	5.84–7.21	114–143
<i>PKP-SKKS</i>	7.97–12.66	126–180
<i>PKP-SS</i>	20.05–27.66	136–180
<i>PKIKP-PKS</i>	3.35–3.60	126–141

mechanism. When the comparison between the phase attributes and the values for a model source is made, the weights and required level of fit for different attributes need to be flexible. This means that there will be of the order of $6N_p \times 360$ features to be individually considered, that is, typically, several thousand cases. In addition, any modification to the system (e.g., to increase the sampling of the model space) will impose a reconsideration of an already complex algorithm.

Since the algorithm of exhaustive matching is inappropriate, we turn to a heuristic method and try to exploit the rules of thumb employed by expert human analysts.

Expert System Techniques

When the seismological expertise in interpreting seismograms (see, e.g., Richter, 1958; Kulhánek, 1990) is summarized into a set of rules, the major problem lies in how to schedule the application of the various rules to produce a solution, which is also called the problem of “knowledge application.” In the field of expert systems, typical examples of the control mechanism for reasoning are “backward chaining” and “forward chaining” (see, e.g., Jackson, 1990). Backward chaining entails taking a conclusion first, then seeking evidence to support it. Forward chaining works in the reverse direction, the system tries to find a rule whose conditions are now satisfied by the information.

In the problem of seismogram interpretation at a single station, we are frequently faced with insufficient evidence (obscured phases) and insufficient expert knowledge (there are cases when even an experienced analyst could fail to produce any solution), while the choices for the solution are many (360, see previous section). In such a context, using the backward chaining method, the reasoning procedure could be bogged down in endless searching; while using the forward chaining method, the reasoning procedure might not be able to progress to any solution (due to insufficient evidence).

The difficulty lies in the fact that the unknown factors (origin time, distance, depth, and the identities of detections) can jointly produce or explain an observation of seismic event. As an example, in order to identify phases, we need information about the epicentral distance. On the other hand,

to obtain the estimation of epicentral distance from a single record, we have to identify at least two phases—we can then usually use the differential time between the identified phases to determine the distance.

In order to make the process tractable, we make a set of hypotheses about the identities of two observed phases. Thus, we select the first *P* phase and the first *S* phase in a sequence of detections as the key phases. The phase character assignment will be based on measures such as \mathcal{P} and \mathcal{S} , as discussed in Tong and Kennett (1995). We then try to associate the pair of phases with the most commonly observed combinations of *P* and *S* phases for different epicentral distance ranges, as listed in Table 1. The dependence of the differential times for these *P-S* combinations on epicentral distance and depth are illustrated in Figure 1. We can see that most of these pairs of phases have sufficient variation in differential times to provide some constraint on range and depth. However, there will often be ambiguity in the appropriate combination of epicentral distance and depth associated with a particular differential time that can only be resolved by the matching of other phases in the record.

We will employ this set of nine *P-S* pairs as the first stage of hypothesis testing as indicated in Figure 2, which represents the classification hierarchy. Since there is some degree of overlap between the operative distance intervals for a number of the *P-S* pairs, the tree structure in Figure 2 is not a strict classification. The number of leaves on the tree is actually more than 360, while the total number of different pattern classes that they represent remains 360.

In the following section, we give a brief description about a new strategy termed an “assumption tree” based on the structure illustrated in Figure 2.

An Assumption Tree

The assumption tree is a multi-branch tree that is proposed here to guide the application of expert knowledge when there are a number of interacting factors to be identified. The tree grows its branches as a set of assumptions and are tested, and so we call it an assumption tree. At each level of the assumption tree, a set of assumptions about a single factor are tested. The set of stems springing from the same branching point will cover the full range of possible hypotheses connected with this factor. Every branching point (node) in the tree has an associated package of information. At the root level, the information consists of just the observed data. Every new node in the assumption tree will inherit all the information of its parent node and include extra information. The new information includes the premise that has just been made to produce this node and the inferences based on the particular assumption drawn from the summary of expert knowledge held in a separate knowledge bank.

The information package at a node is examined to look for any contradictory results. If any contradiction is found, the branching associated with the current node will be terminated. Otherwise, the branching process will continue

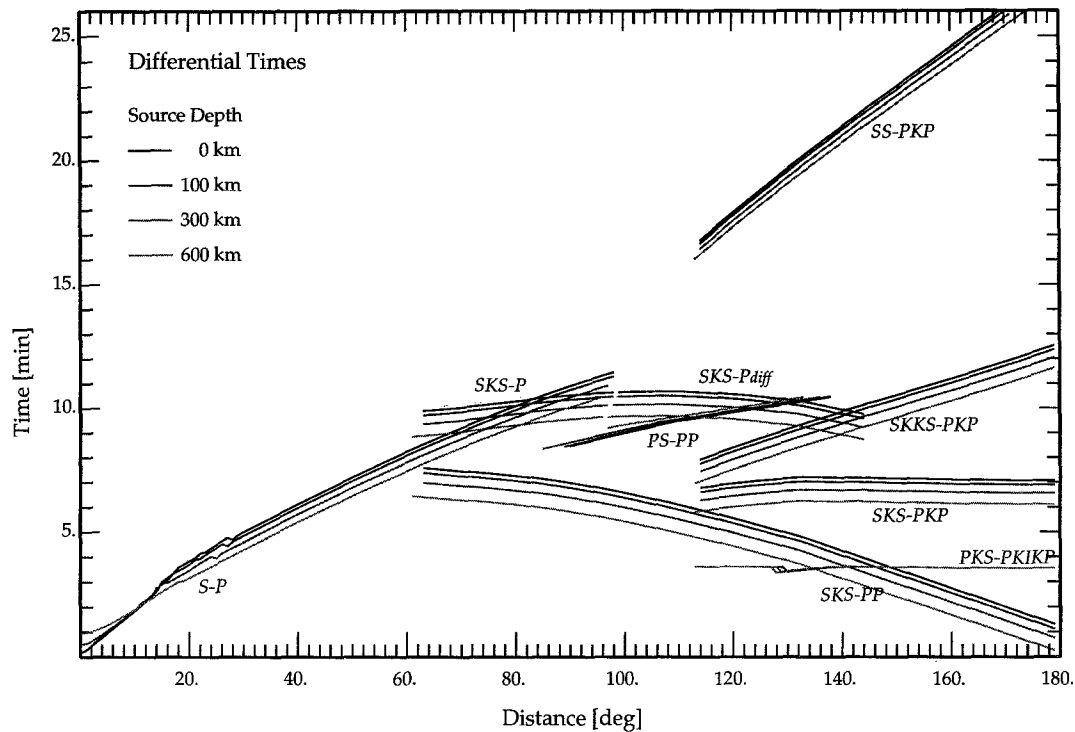


Figure 1. The distance and depth dependence of P - S differential times.

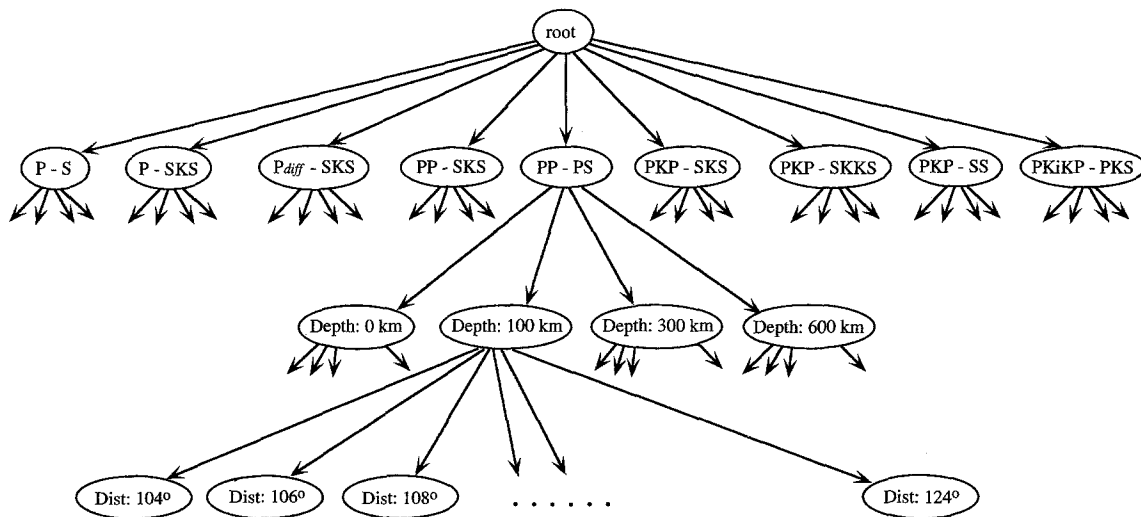


Figure 2. Seismic event classification hierarchy.

from the node and allow the evaluation of further hypotheses. At any node, the new information introduced from the knowledge bank may limit the choices for other currently undetermined factors. As a result, the pattern of branching at a node will depend on the current information package. When no further assumptions are to be tested, we reach a leaf of the tree representing a solution based on a set of hypotheses about the factors describing the data.

Automatic Seismic Event Recognition and Later Phase Identification

Phase Detection and Feature Extraction

We have employed the procedures described by Tong and Kennett (1995) for phase detection and feature extraction. In the application of these techniques to the problem of event identification, we try to ensure that the maximum

opportunity is made to extract the relevant phases. The first stage uses a detection system based on a complexity measure (see Tong and Kennett, 1995) to detect the approximate position for a phase package. Then the original seismogram is divided into a number of segments that contain different phase packages for each of which the local frequency can be estimated, as described by Tong and Kennett (1995). Each seismogram segment is then filtered into a high-frequency trace and a low-frequency trace with filter parameters based on the local phase frequency. Finally, a set of detections are made on both the low-pass and high-pass filtered traces using adaptive STA/LTA detectors that combine energy measures on one or more components of ground motion with a local frequency analysis. To simplify the input for event identification, we choose to detect only the first phase where there are complicated phase packages.

In order to be able to gain reliable information about a phase, we want not only to detect the arrival but also to characterize the phase in terms of a phase vector comprising a summary of the information in the arrival. By filtering segment by segment, we avoid problems associated with a wide variation between the frequencies on the three components, and by using a common filtering for all three components, we improve the reliability of the phase vector information.

Assembling an Event

An event on a seismic record that is being subjected to real-time analysis can be recognized as a number of *P* phases followed by at least one *S* phase and possibly a number of other *P*- or *S*-type phases within a 30-min window. With this working definition of an event, a set of detected phases (and their associated parameters) can be assembled into a list that will be used as the basis for event classification and further phase identification.

Seismic Event Interpretation

The system of event processing is based on an assumption tree with a similar structure to that illustrated in Figure 2. The assumption tree grows at run time as we test a set of hypotheses about the observed seismic phases; branches that lead to contradictory results will be terminated and not grow to leaves representing a viable event classification.

Stage 1—Choice of *P*–*S* Pair. The initial information package comprises the set of information on the detected phases at the root node. Nine stems are then grown to child nodes representing different assumptions about the identity of the combination of key *P* and key *S* phases (as indicated in Fig. 2). The consequences of these assumptions are then tested against the expert knowledge held in the separate knowledge base.

If the assumption about the *P*–*S*-phase pair is true, the event must lie within the distance range for this combination (see Table 1), and the differential time between the first *P* and the first *S* must lie inside the possible *P*–*S* time interval.

The possible distance range is thereby reduced from 0° to 180° to the distance range appropriate to the particular choice of phases. The observed *P* to *S* differential time can then be compared with the expected range for this phase combination. If the observed differential time does not lie inside the expected range, we have a contradiction of the assumption about the phase pair, and the branch of the assumption tree is terminated. The phase pairs in Table 1 have some overlap in differential times so that, at most, five possible assumptions about the *P*–*S*-phase character can survive this test.

At this stage, we can apply additional seismological information to test the different premises about the *P*–*S* pair. The properties that have been employed are tabulated in Table 2.

Stage 2—The Depth of the Event. At this stage, every surviving first-generation node grows four new stems to child nodes (see Fig. 2), where each child node is associated with a different assumption about source depth. Based on the assumptions about the identity of the *P* and *S* phases and depth and the observed *P*–*S* differential time, we are able to isolate a single possible epicentral distance. Nine differential-time tables (one for one of the *P*–*S* pairs) connect the *P* to *S* time interval with epicentral distance. A search in the appropriate table will return the distance that corresponds to the smallest difference between the theoretical and the observed differential time.

Since we find the distance by table search, we do not implement the third-generation assumptions for distance (indicated in Fig. 2), and as there are no more assumptions to test, we have reached a leaf of the assumption tree.

With the estimates of the depth and distance, we can compare the patterns of observed phases with those expected for the nine different assumptions on the *P*–*S* pairs. As de-

Table 2
P–*S* Pairs and Related Feature Properties for Earthquakes

<i>P</i> – <i>S</i> Identity	Expected Features
<i>P</i> – <i>S</i>	The frequency of <i>S</i> should be lower than the frequency of <i>P</i> ; <i>P</i> should not be too steep (as <i>PKP</i>); there should not be any low-frequency <i>S</i> phase before the <i>S</i> .
<i>P</i> – <i>SKS</i>	<i>SKS</i> should not be a high-frequency phase.
<i>P_{diff}</i> – <i>SKS</i>	<i>P</i> has a smaller amplitude compared with <i>SKS</i> .
<i>PP</i> – <i>SKS</i>	<i>PP</i> : long-period and shallow incidence.
<i>PP</i> – <i>PS</i>	<i>PP</i> : long-period and shallow incidence.
<i>PKP</i> – <i>SKS</i>	<i>PKP</i> : short-period and steep incidence.
<i>PKP</i> – <i>SKKS</i>	<i>PKP</i> : short-period and steep incidence.
<i>PKP</i> – <i>SS</i>	<i>PKP</i> : short-period and steep incidence; <i>SS</i> : long period.
<i>PKIKP</i> – <i>PKS</i>	<i>PKIKP</i> : short-period and steep incidence.

The specification “steep” or “shallow” depends on the apparent angle of incidence; the smaller the incidence angle, the steeper. *PKP* and *PKKP* are observed for very distant events for which the incidence angle is very steep, normally less than 16°. Thus, an observed phase should have an incidence angle smaller than 16° to be matched with *PKP* or *PKKP*. The value of 16° has been obtained by trial and error and is subject to further adjustment.

Table 3

Expected Phases for Choice of P - S Pairs, for a Certain Distance and Depth Ranges

P - S Identity	Other Expected Phases
P - S	Deep event: depth phases pP , sP ; 25°–65°: PcP , ScP , PcS ; 12°–39°: ScS ; 31°–79°: PP .
P - SKS	for all: PP ; 95°–99°: $PKKP$.
P_{diff} - SKS	for all: PP ; 105°–125°: $PKKP$, PS .
PP - SKS	105°–125°: $PKKP$, PS ; 95°–99°: $PKKP$; for all: SS .
PP - PS	for all: $PKKP$, SKS , $SKKS$.
PKP - SKS	Deep event: depth phases $pPKP$, $sPKP$; for all: PP ; 129°–143°: $SKKS$; 115°–125°: $PKKP$, PS ; 25°–143°: PKS ; 114°–131°: SS .
PKP - $SKKS$	Deep event: depth phases $pPKP$, $sPKP$; for all: PP ; 125°–139°: PKS ; 137°–180°: SS ; 159°–180°: PKP_{ab} .
PKP - SS	Deep event: depth phases $pPKP$, $sPKP$; 141°–180°: PP ; 159°–180°: PKP_{ab} .
$PKIKP$ - PKS	Deep event: depth phases $pPKIKP$, $sPKIKP$; for all: PP , $PKKP$, $SKKP$; 137°–143°: SS .

scribed above, at most, five models for the P - S pair can survive from the first generation. Once four possible depths are included, there would be, at most, 20 live nodes in the second generation. Since 20 is not a large number, the analysis procedure accepts all the second-generation nodes as possible interpretations, and then we try to assess the likelihood of each interpretation rather than look for contradictions.

Identification of Other Frequently Observed Phases. So far, we have used seismological information on the likely character of the two main phases (the first P phase and the first S phase) that are required to form an event pattern. For a choice of source distance and depth, we can make use of a set of “other frequently observed phases,” as summarized in Table 3; the arrival times and other properties can then be calculated.

Then, by comparison between the sequence of observed phases and the phase predictions for the assumed distance and depth, we may be able to identify phases, and further, the presence of expected phases can provide support for the assumed depth and distance.

The specific choice of the P - S pair and the other expected phases constitute the information set to be compared with the observed set of phases. We therefore need to match the observations against the expectations for the proposed distance and depth. For each expected phase, we have an arrival time determined by the phase identity; each detected phase has a set of observed phase attributes that includes the arrival time. The first P and first S phase in the observed data have already been associated with two phases; we can

Table 4

Expected Properties of Other Frequently Observed Phases

Phase Identity	Expected Features
pP , sP	pP , sP are similar to P both in frequency and incidence.
PcP , ScP	PcP , ScP are high-frequency P phases, which is steeper than P .
PP	PP is shallower than P and lower frequency.
$PKKP$	$PKKP$ is a steep and high-frequency P phase.
PKP	PKP is a steep P phase.
ScS	ScS : high-frequency steep S phase, strong on tangential component.
PcS	PcS : high-frequency steep S phase.
PS	PS is shallow and low-frequency S phase.
SS	SS is a low-frequency S phase.
SKS , $SKKS$	SKS , $SKKS$ cannot be high frequency.

therefore establish a simple time mapping between the observed data and the expected arrivals. We search among the observed arrivals for candidates for an expected phase whose arrival time lies within a predefined tolerance of the expected time. The observed attributes of the phase are then compared with a summary of phase properties in the knowledge base (see Table 4). If there is no match, then that hypothesis for the phase identification is rejected, but if more than one phase match is possible, the one with the smallest time difference between observed and expected times is selected.

Estimation of Likelihood of Hypotheses. In order to estimate the likelihood of any particular interpretation, we consider both the observed data and the expectation for a particular combination of P - S pair, distance and depth. When we compare the two sets of phases, we have three different cases: matched phases, unmatched phases in the observed data, and unmatched phases in the expected data.

For every matched phase, a time error e_i is used to represent the mismatch between the corresponding observed and expected phases. We set

$$e_i = \frac{|t_0 - t_e|}{t_s - t_p}, \quad (1)$$

where t_0 is the phase time in the observed data, t_e is the phase time for the expected phase, t_s is the observed time of the key S phase, and t_p is the observed time of the key P phase. For every unmatched phase (both in the observed data and the expected phase list), we set $e_i = 1$.

To provide an overall measure of the quality of the match between the observed phases and those expected for the particular distance and depth, we weigh each observed phase with its amplitude (i.e., we place more weight on large arrivals). For those expected phases that have not been successfully matched with any observed phase, the smallest amplitude of the matched components is applied as the weight. We denote the weight for each phase as w_i , and then we define the misfit measure δ as

Table 5
Events Used for Illustration

Event	Year	Day	Time	Station	Distance (°)	Depth (km)	Azimuth (°)	M_s
A	1994	281	21:45:55.8	SC03	15.4	17	175	6.4
B	1994	220	21:15:36.6	SC01	52.0	122	133	6.0
C	1994	244	15:20:18.8	SC08	113.8	10	260	6.6
D	1994	160	00:46:55.9	SC09	137.1	631	213	7.0

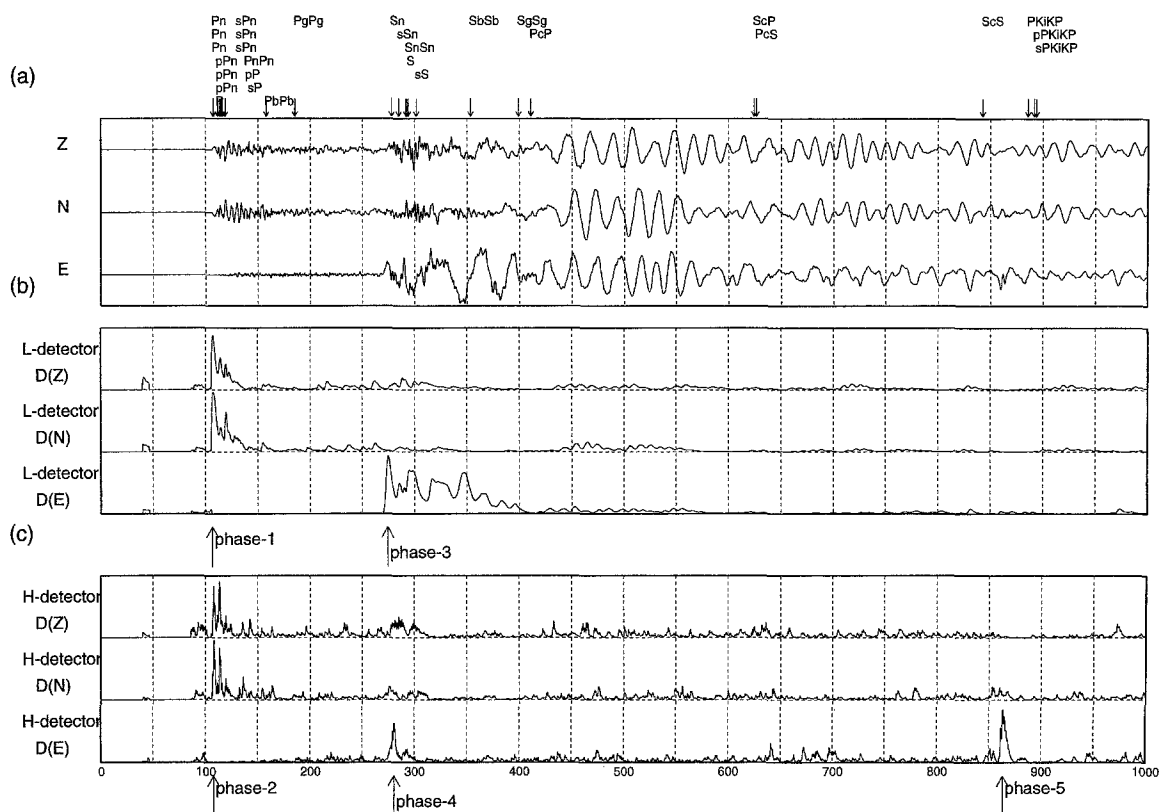


Figure 3. Phase detection for event A (for event details, see Table 5)—Halmahera (Indonesia). (a) The original broadband record is displayed together with the arrival times for the major phases predicted from the IASP91 travel-time tables as a reference for the phase interpretation process. (b) The STA/LTA detectors generated for the three low-pass filtered components (Z, N, E). The arrows indicate where phase detections have been made. (c) The STA/LTA detectors on the three high-pass filtered traces are displayed, along with the arrows that indicate phase detections. The numbered sequence of the phase detections in (b) and (c) is in order of arrival time.

$$\delta = \frac{\sum_{i=1}^n e_i w_i}{\sum_{i=1}^n w_i}, \quad (2)$$

where the sum is taken over n , which is the total number of matched and unmatched phases.

Other Details of the Implementation. We take a pragmatic approach to the selection of phases for the key P – S pair since we have used the most commonly observed phases; rather than insist on the “first” arrival of particular type, we look for a prominent arrival, especially for S . After the detection

of a first P phase, we choose the first S phase encountered in the processing scheme as the candidate for the key S phase. However, if a second S phase occurs within 5 min without any intervening phase and is larger, it will replace the first candidate.

The detection procedures use adaptive high- and low-pass filtered traces in phase characterization (see Tong, 1995). The low-passed seismogram gives a better definition of phase attributes than the corresponding high-passed seismogram. Therefore, when we select the two key phases, we choose the detections from the low-frequency trace.

We use the azimuth estimate for the key P phase as the

azimuth of the whole event. Since the estimated azimuth actually implies two possible directions (with a 180° difference in between), both the possible values for azimuth are given in the interpretation result. For very distant events, from 100° onward, the first P arrival is usually somewhat weak, and the resulting azimuth estimate is somewhat doubtful.

Examples of Event Identification

The event-recognition system described in the previous section has been applied to a wide range of observed seismic records with considerable success. We have selected a set of four examples that represent events from different distances and depths.

We use seismic records from portable broadband instruments deployed in the Northern Territory of Australia. The data were recorded on Reftek 72A-07DAT recorders (24-bit resolution) with Güralp CMG-3ESP seismometers (flat to ground velocity from 0.03 to 30.0 Hz) with GPS timing. The sampling rate was 25 samples per second. Such data are of good quality but not quite of observatory standard, and so the data represent a useful test for both event recognition and phase identification.

The four events we use for illustration are listed in Table 5. We note that the event locations as listed in Table 5 have been determined using information from many stations with a broad azimuthal coverage, whereas our estimates are based on just a single, three-component record. The time listed in the table is the beginning time of the record from which the automatic system is applied.

For each event, we present a figure with three display panels, representing the original seismograms and two sets of phase detection curves derived from the low- and high-pass filtered records, together with the phase detections employed in the analysis. For each phase detection, six phase features are extracted, as described by Tong and Kennett (1995), and these are summarized in a table for the event along with the processing sequence leading to the characterization of the event and the phases. The procedure is organized in terms of the four major steps in the analysis:

- Step 1: Identification of the key P and S phases.
- Step 2: Comparison of the differential time between the key phases with the expected times for P – S pairs.
- Step 3: Checking of expected P – S pair phase attributes.
- Step 4: Matching of observed phases against expected patterns for different models of event character.

The final step is selection of the combination of distance and depth that leads to the least misfit between the observed and expected phases.

Event A—Halmahera (Indonesia)

The event shown in Figure 3 is a regional event at a distance of 15° . In panel (a), the original seismic record is dominated by Pn and Sn packages. Later phases are mixed

with surface waves. Since the P arrival shows little energy on the E component, the seismogram is almost naturally polarized; that is, the E component is the tangential component that reveals SH -wave arrivals.

In panels (b) and (c), the Pn and Sn arrivals are detected on both high-frequency and low-frequency sets. Since ScS is strong on the tangential component, the high-frequency set E component clearly reveals the ScS arrival (see [phase-5] in panel c). The detections [phase-1] and [phase-2] occur at almost the same time on the two differently filtered sets, so also do [phase-3] and [phase-4]. The automatic system uses the attributes for the phase on the low-frequency set from each pair; it therefore chooses [phase-1] as the key P phase and [phase-3] as the key S phase for the differential time analysis. Table 6 summarizes the progress of the event interpretation scheme. After the test of the differential times, only the $\{P, S\}$ pair survives. This P – S model also survived the test on the expected P – S features. Therefore, there are only four nodes in the second generation corresponding to different interpretation of the event depth.

For every choice of seismic source interpretation in step 4 of Table 6, the distance range and P – S pair lead us to expect the arrival of the phase ScS . For the first case, with a surface source (depth 0 km), the expected time for ScS corresponds to the observed [phase-5], and the expected phase features for ScS are matched with the attributes of [phase-5]. Therefore, in addition to the two key phases, a further phase is identified. For each of the other three cases, the expected time range for ScS does not cover any observed phases, as a result, the expected phase ScS is not found, and the misfit measures are larger than that for a surface source.

When we compare the best interpretation (distance 14° , depth 0 km, azimuth 180°) with the event specification given in Table 5, we can see that the estimations of distance, depth, and azimuth are very close to the true values. At the same time, the automatic system has correctly identified three phases: [phase-1] as P , [phase-3] as S , and [phase-5] as ScS .

Event B—Myanmar (Burma)

This is a typical teleseismic event that is dominated by clear P , S , and core reflections. In Figures 4b and 4c, we see that P and S arrivals are detected on both the high-frequency and the low-frequency set. Other detections are [phase-3] for PcP and [phase-4] for PcS . In Figure 4b, the detection traces also indicate the presence of pP , PP , and sS arrivals. However, these detections are not flagged with an arrow because they lie below a predefined threshold. The threshold has been set to a common value for all the illustrations and appears to be a little too high for this unusually clear event.

The phases identifications above can be made because we have external information that is not available to the automatic event-recognition procedure. Now, let us see how the automatic system determines the event location and identifies the phase detections. The automatic system chooses the low-frequency detections [phase-2] as the key P phase and [phase-6] the key S phase. The subsequent steps in the

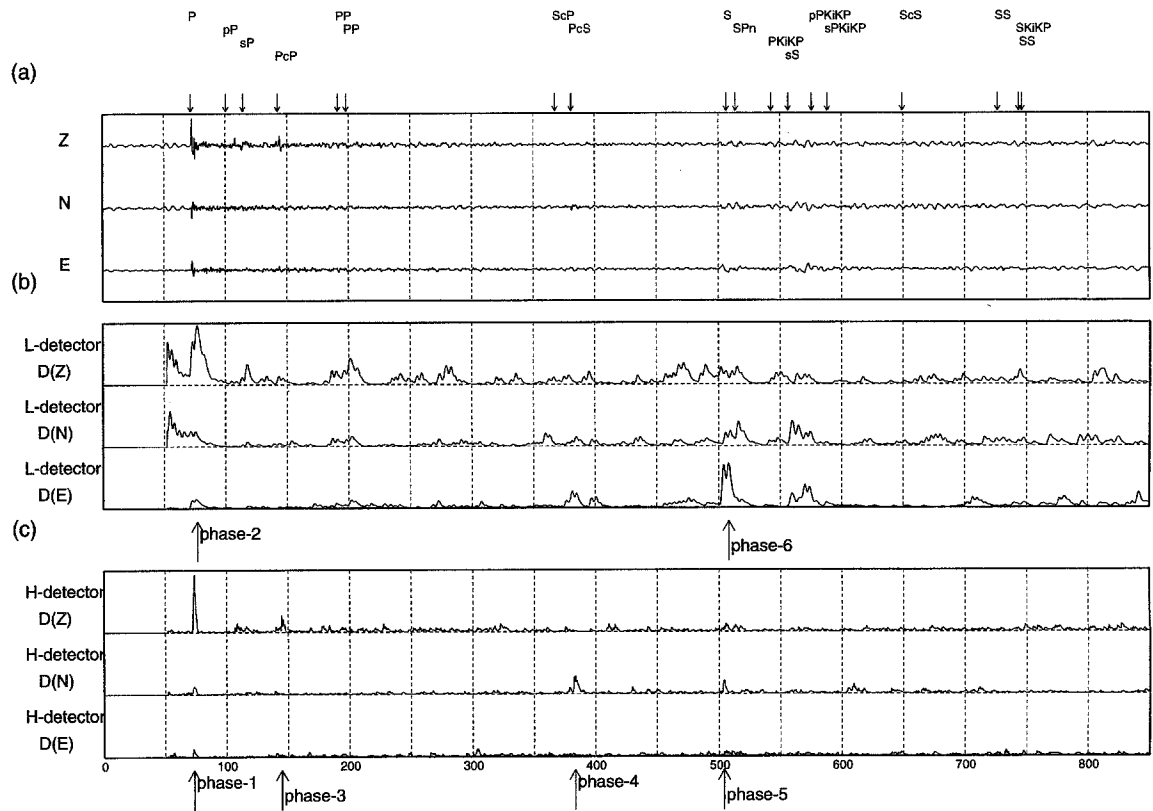


Figure 4. Phase detection for event B—Myanmar (Burma). For explanation, see Figure 3.

Table 6
Processing Sequence for Event A

Input: Phase	1	2	3	4	5
t_i (sec)	107.2	108.0	275.0	280.6	863.4
a_i	0.060	0.030	0.175	0.344	0.102
v_i (Hz)	0.136	0.568	0.057	0.431	0.321
ϕ_i (°)	180.3	180.5	85.7	250.2	263.0
ψ_i (°)	50.2	49.6	9.2	11.9	1.4
c_i	<i>P</i>	<i>P</i>	<i>S</i>	<i>S</i>	<i>S</i>

Processing:
Step 1: choice of key phases
Key *P*: phase 1; key *S*: phase 3

Step 2: test on differential time
167.8 sec—surviving models: *P*–*S*

Step 3: *P*–*S* pair feature match
Surviving models: *P*–*S*

Step 4: <i>P</i> – <i>S</i> Pair	Depth	Dist.	Azim.	Event Pattern Match Identified Phases	Misfit
<i>PS</i>	0	14	180 or 0	phase 1: <i>P</i> ; phase 3: <i>S</i> ; phase 5: <i>ScS</i>	0.0033
	100	16	180 or 0	phase 1: <i>P</i> ; phase 3: <i>S</i>	0.3333
	300	16	180 or 0	phase 1: <i>P</i> ; phase 3: <i>S</i>	0.3333
	600	16	180 or 0	phase 1: <i>P</i> ; phase 3: <i>S</i>	0.3333

event interpretation scheme are summarized in Table 7. After the test on the differential time between the key *P* and *S* phases, the three pairs {*P*, *S*}, {*PP*, *SKS*}, and {*PKP*, *SKS*} have survived. Since the apparent angle of incidence of the key *P* phase is 27.8°, which is not as steep as required for *PKP* (see the note in Table 2), the choice of {*PKP*, *SKS*} fails in the test on *P*–*S* features. The other two *P*–*S* combinations survive this feature test, and so there are eight nodes in the second generation in the assumption tree corresponding to different combinations of distance and depth as well as phase choice.

The smallest measure of misfit occurs for the second choice in step 4 of Table 7, which corresponds to an interpretation of the key phases as *P* and *S*. The estimates of the event parameters are epicentral distance 52°, depth 100 km, and azimuth 316° (which also implies 136°), which matches the information provided in Table 5 very well.

Event C—Off the Coast of Northern California

Event C is at a much greater distance and so is in the core shadow for *P*. The expected first arrival, diffracted *P*, is so weak that it cannot be seen on the seismogram (Fig. 5a). Core phases (e.g., *PKP* and *SKS*) and surface multiples (e.g., *PP* and *PS*) are the main features on seismic records at this distance range.

The set of detections displayed in Figures 6b and 6c

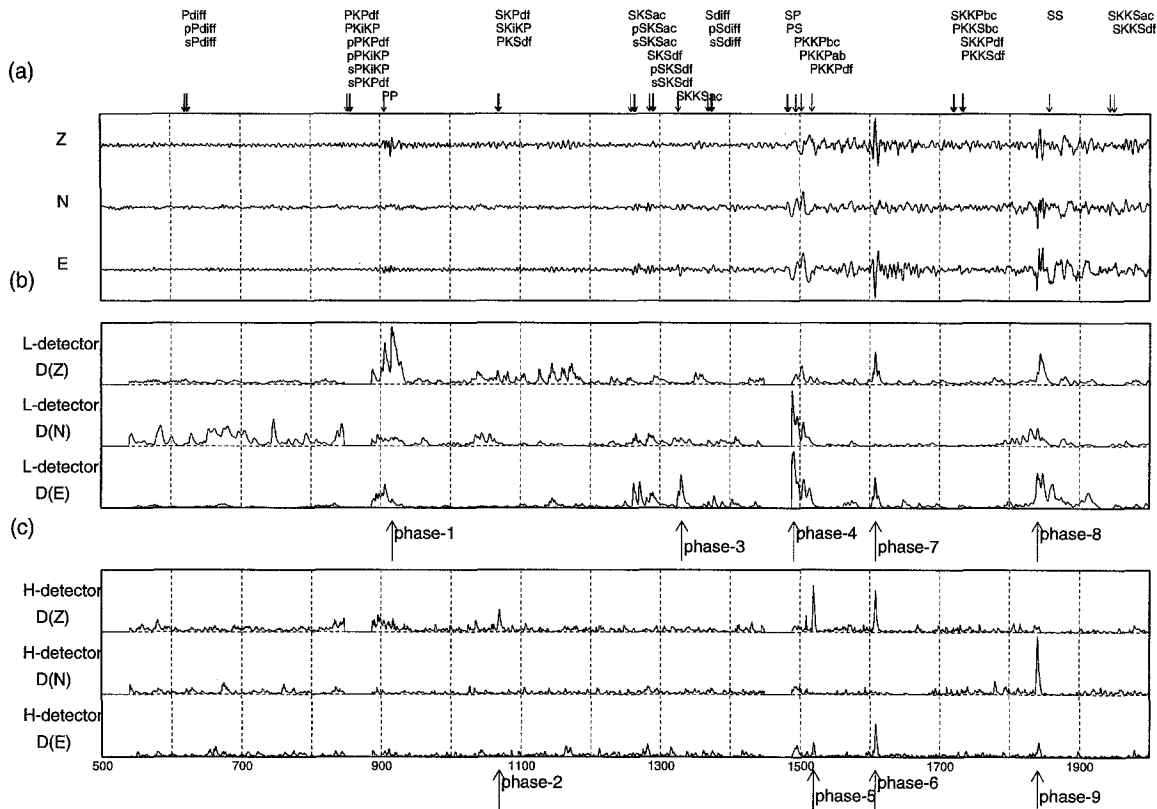


Figure 5. Phase detection for event C—off the coast of northern California. For explanation, see Figure 3.

pull out all the major arrivals on the seismogram. We notice that the detection [phase-7] and another high-frequency S arrival [phase-9] have no obvious correlation with the phases expected for this event and may well come from some different source.

The detection on the low-frequency set [phase-1] is selected as the key P phase. Note that since [phase-4] is bigger than [phase-3] and there is no detection in between, [phase-4] is selected as the key S phase in place of [phase-3]. The progress of the automated event-recognition system is presented in Table 8. In the first generation in the assumption tree, five sets of P – S pairs survived the test on the differential time between the key phases. For the choice $\{P, S\}$, we do not expect to see another low-frequency S phase before the key S phase; therefore, this choice is unsuitable. In the case of $\{P_{diff}, SKS\}$, the amplitude of the key P phase is expected to be much smaller than the key S phase. However, the observed behavior is that the key P phase is almost comparable with the key S phase; therefore, this choice is also rejected. For the choice $\{PKP, SKS\}$, the angle of incidence of the key P phase is not suitable for a PKP , as also occurred for event B. Finally, the surviving phase pairs at the first-generation nodes are $\{P, SKS\}$ and $\{PP, PS\}$.

In step 4 of Table 8, we note that the most likely interpretations are for sources at either the surface or 100 km in depth, at a distance of 112° with $\{PP, PS\}$ identified as the

pair of key phases. The level of misfit is very close for the two cases. When we compare the event-parameter estimates with the event specification in Table 5, we can see that the distance is very well determined; the estimate of depth is reasonably good, while the estimated azimuth is far from the true value because of the distance. However, it is interesting to see that the azimuth 263° for [phase-5] in Table 8 (identified as $PKKP$) is very close to the true azimuth 260° .

The larger misfit for event C compared with the previous cases is caused by the uncorrelated big arrivals (i.e., [phase-7] and [phase-9]), which most likely come from an overlapping event (note that there is no obvious association with the phases predicted from the IASP91 travel times).

Event D—Northern Bolivia

The epicentral distance is even larger for this event, and the seismogram is now dominated by core phases, such as PKP , $PKKP$, SKS , $SKKS$, etc. Figure 6a shows two complicated packages of core arrivals in the early part of the broadband record.

The phase detection system is designed to pick out the first onset of a phase package. Therefore, the detections [phase-2] and [phase-4] are indicated at the beginning of the two packages. From the low-frequency detections, [phase-2] is selected as the key P phase and [phase-6] as the key S phase. As before, we can follow the progress of the automatic system

Table 7
Processing Sequences for Event B

Input: Phase	1	2	3	4	5	6
t_i (sec)	73.5	76.8	144.9	383.5	504.2	508.1
a_i	0.590	0.059	0.151	0.057	0.040	0.052
v_i (Hz)	0.893	0.140	0.446	0.431	0.625	0.128
ϕ_i (°)	140.7	316.4	131.5	10.8	170.3	286.8
ψ_i (°)	27.0	27.8	25.4	9.5	33.6	5.1
c_i	<i>P</i>	<i>P</i>	<i>P</i>	<i>S</i>	<i>S</i>	<i>S</i>

Processing:

Step 1: choice of key phases
Key *P*: phase 2; key *S*: phase 6

Step 2: test on differential time
431.3 sec—surviving models: *P-S*, *PP-SKS*, *PKP-SKS*

Step 3: *P-S* pair feature match
Surviving models: *P-S*, *PP-SKS*

Step 4:

<i>P-S</i> Pair	Depth	Dist.	Azim.	Event Pattern Match Identified Phases	Misfit
<i>PS</i>	0	50	316 or 136	phase 2: <i>P</i> ; phase 3: <i>PcP</i> ; phase 6: <i>S</i>	0.00196
	100	52	316 or 136	phase 2: <i>P</i> ; phase 3: <i>PcP</i> ; phase 4: <i>PcS</i> ; phase 6: <i>S</i>	0.00015
	300	54	316 or 136	phase 2: <i>P</i> ; phase 3: <i>pP</i> ; phase 4: <i>PcS</i> ; phase 6: <i>S</i>	0.00031
	600	58	316 or 136	phase 2: <i>P</i> ; phase 3: <i>PcP</i> ; phase 4: <i>PcS</i> ; phase 6: <i>S</i>	0.00105
					0.33333
<i>PP SKS</i>	0	84	316 or 136	phase 2: <i>PP</i> ; phase 6: <i>SKS</i>	0.33333
	100	82	316 or 136	phase 2: <i>PP</i> ; phase 6: <i>SKS</i>	0.33333
	300	82	316 or 136	phase 2: <i>PP</i> ; phase 6: <i>SKS</i>	0.33333
	600	82	316 or 136	phase 2: <i>PP</i> ; phase 6: <i>SKS</i>	0.33333

in Table 9. Three choices for the pair of key phases survive the test of differential time. Since the incidence angle for [phase-2] is 9.78°, which is too steep to be a *P* phase in the choice of the {*P*, *S*} pair, the branch corresponding to {*P*, *S*} is terminated. Similarly, the incidence angle and the frequency of [phase-2] is not suitable for a *PP* in the {*PP*, *SKS*} pair. Therefore, only one choice of the *P-S*-phase pair model {*PKP*, *SKS*} is left after the *P-S* feature test.

The most likely interpretation identified in step 4 of Table 9 is the last choice of a very deep event at a distance of 128°. Once again, the large epicentral distance means that the azimuth of the first *P* phase is not a reliable estimate. When we compare with the event specification in Table 5, the estimated depth is quite accurate. However, the estimated distance is 9° away from the true value, which is not as good as in the previous examples. This difference is actually caused by the procedure described in a previous section (Stage 2—The Depth of the Event) because the mapping between differential time and distance is not one-to-one for the case of *SKS-PKP* (see Fig. 1). When only one choice of

distance is made from the mapping (differential time to distance), the true distance can be missed.

Performance of the Event-Recognition System

As applied to a wide range of seismic events, in every case, the automatic system can successfully restrict attention to a limited number of possible interpretations, and then a quantitative assessment can be made by using the misfit measure between the observed and expected phases. The misfit measure performs very well in selecting the most likely interpretation. The more phases detected and identified while recognizing a seismic event, the better discrimination (between the possible solutions) the misfit measure can give. In most cases, the estimated distances are within the error of 7%. And the estimated azimuths are within the error of 5° for distances less than 100°. For an event at a greater distance (beyond 100°), the azimuth could not be well determined by only analyzing the single-site data.

In the current procedure, there is a tendency for the depth to be overestimated. For shallow sources, this can occur when a feature on the record is associated with a near-source reflection (e.g., *pP* and *sP*) since these cannot be correlated with a surface source. Because of the sparse sampling in depth, the procedure is forced to make a decision between 0 and 100 km. Complexity in the source pulse (in the case of these two examples, the interference from the closely following *PcP*) can be mistaken for a depth phase and hence suggest a deeper origin than appropriate. Also, in order to match the observed differential time with the 2° sampling of the field, it may be possible to improve the match by adjusting the depth. Further, the IASP91 travel-time tables have been used in the analysis for which *S* is now recognized to be a little slow (Kennett *et al.*, 1995), and this can be compensated by increasing the depth of the source.

The various tests indicate that, in general, the performance of the automated analysis procedure is good, provided that it can be supplied with good quality information on detected phases and their attributes. The application of the automated procedures for the same event recorded across the broad network of portable stations in northern Australia, from which the examples above have been drawn, has shown that very consistent results can be obtained under varying local conditions and that constraints on location can be imposed to within a 4° cap. Such a preliminary location estimate and its associated depth would be very valuable in minimizing processing time for a global multi-station network.

Discussion

Extensions of the Processing Scheme

The present event-recognition system has been designed to be simple and robust so that we can concentrate on the processing strategy. We recognize that, at present, we have not included the full range of available seismological information on the expected character of the seismic wave field.

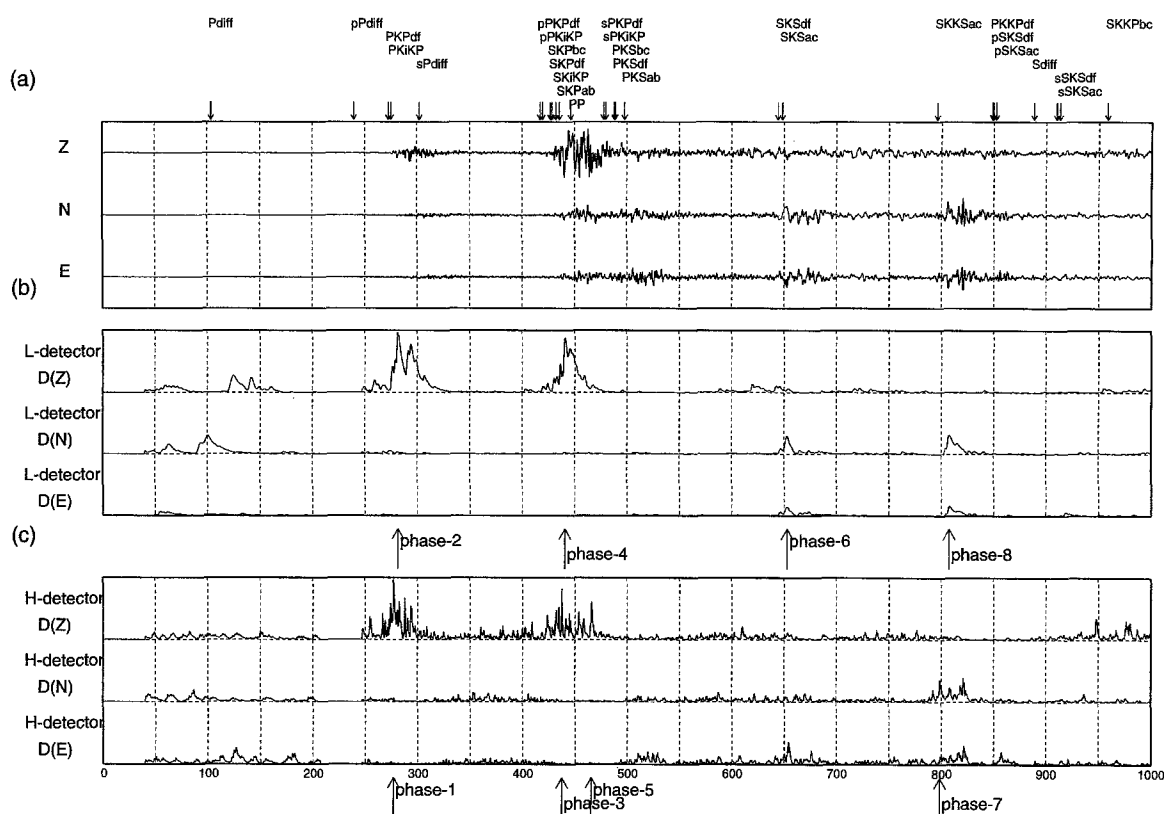


Figure 6. Phase detection for event D—northern Bolivia. For explanation, see Figure 3.

For example, for some source orientations and azimuths, it is possible for the depth phases, for example, pP , sP , and pS , sS , to be substantially larger than the equivalent direct phases, and so a depth phase could be picked as a key phase with consequent distortion of the present phase-identification process. Fortunately, an important feature of the assumption tree approach is that it is relatively easy to make revisions and improvements in the future.

First, the seismological expertise for the nature and properties of commonly observed phases (Table 3 and 4) is kept separately from the algorithm. Thus, the addition or modification of the summary of this seismological expertise will not affect the algorithm.

More significantly, the information required for constructing the assumption tree is also held separated from the algorithm itself. For example, the growth of the stems to the first-generation nodes, representing the set of P – S pairs, is implemented by reading a separate list of P – S choices (Table 1). The automatic system will construct a first-generation node for every item in this P – S list. Thus, the inclusion of further P – S pairs can be readily accomplished by adding more items to the P – S list. A similar approach can be applied to the second generation, to revise or extend the possible source depths.

The sampling in distance could be increased by employing a denser travel-time table for each depth. The current

number of depth/distance combinations (360) has been chosen to allow rapid testing for a possible real-time environment while still providing a useful definition of the nature of an event for preliminary location. However, the number of classes could be increased without a very large processing overhead to provide an increased density of states.

Further Developments

Refinement of Source Distance and Depth. As previously described, the source distance is obtained by inference from the P – S model and the differential time interval. For a typical combination of P – S phase pairs, the differential time has an almost linear dependence on the source distance. This property is exploited in the search to find the closest match to the observed differential time.

There are two factors that need to be taken into consideration: first, we have to expect some differences between the observed differential time and that for an Earth model such as IASP91, and so we need to allow for some tolerance when matching times. Second, the mapping between differential time and distance is not always one-to-one, as can be seen for the $\{PKP, SKS\}$ and $\{PKIKP, PKS\}$ pairs in Figure 1; this occasional problem could be resolved by using a third generation in the assumption tree (Fig. 2) to determine distance.

Table 8
Processing Sequence for Event C

Input: Phase	1	2	3	4	5	6	7	8	9
t_i (sec)	915.9	1068.7	1330.4	1490.9	1519.0	1607.9	1608.0	1840.2	1840.6
a_i	0.148	0.044	0.124	0.176	0.284	0.402	0.523	0.338	0.430
v_i (Hz)	0.208	0.446	0.151	0.056	0.781	0.625	0.152	0.179	0.521
ϕ_i (°)	45.0	228.38	244.8	240.1	262.8	74.7	254.8	235.1	164.4
ψ_i (°)	24.1	5.44	6.6	17.6	30.2	40.9	45.6	26.4	14.6
c_i	<i>P</i>	<i>P</i>	<i>S</i>	<i>S</i>	<i>P</i>	<i>P</i>	<i>P</i>	<i>S</i>	<i>S</i>

Processing:

Step 1: choice of key phases
Key *P*: phase 1; key *S*: phase 4

Step 2: test on differential time
575.0 sec—surviving models: *P-S*, *P-SKS*, *P_{diff}-SKS*, *PP-PS*, *PKP-SKKS*

Step 3: *P-S* pair feature match
Surviving models: *P-SKS*, *PP-PS*

Step 4: <i>P-S</i> Pair	Depth	Dist.	Azim.	Event Pattern Match Identified Phases	Misfit
<i>P SKS</i>	0	82	225 or 45	phase 1: <i>P</i> ; phase 4: <i>SKS</i>	0.8399
	100	82	225 or 45	phase 1: <i>P</i> ; phase 4: <i>SKS</i>	0.8399
	300	82	225 or 45	phase 1: <i>P</i> ; phase 4: <i>SKS</i>	0.8399
	600	92	225 or 45	phase 1: <i>P</i> ; phase 4: <i>SKS</i>	0.8399
<i>PP PS</i>	0	112	225 or 45	phase 1: <i>PP</i> ; phase 3: <i>SKKS</i> ; phase 4: <i>PS</i> ; phase 5: <i>PKKP</i>	0.4662
	100	112	225 or 45	phase 1: <i>PP</i> ; phase 3: <i>SKKS</i> ; phase 4: <i>PS</i> ; phase 5: <i>PKKP</i>	0.4658
	300	110	225 or 45	phase 1: <i>PP</i> ; phase 4: <i>PS</i>	0.7161
	600	108	225 or 45	phase 1: <i>PP</i> ; phase 4: <i>PS</i>	0.7161

The distance and depth estimates can be refined after the main event-recognition process by introducing a local assumption tree to search for the best solution in the distance/depth neighborhood of the postulated event parameters.

Continuous Operation and Overlapping Events. As we have described earlier, the phases corresponding to an event sequence are recognized and separated from the input stream by comparison with a certain expected pattern for an event. The input is a continuous stream of phase detections so that we need also to be able to run the event-interpretation process continuously. The normal procedure in pattern recognition is to use information in time order to separate groups of phases as events, event by event, without going back or reusing the previous data. However, in the case of seismic event recognition, taking into consideration the likelihood of arrivals from different events overlapping in time (as for event C), we would suggest that the unmatched arrivals in previously observed data be combined with the rest of the input stream and reused for recognizing an event sequence, to yield the input to the next event process. In this way, we will be able to separate overlapping events to allow full phase recognition and interpretation.

Acknowledgments

We would like to thank Jan Weekes for sharing her many years experience in the recognition of seismic phases and seismogram interpretation, which have helped to shape the automatic procedures developed in this article.

This work was supported in part by Grant F49620-94-0110 from the U.S. Air Force Office of Scientific Research.

References

- Allen, R. V. (1978). Automatic earthquake recognition and timing from single trace, *Bull. Seism. Soc. Am.* **68**, 1521–1532.
- Bache, T. C., S. R. Bratt, H. J. Swanger, G. W. Beall, and F. K. Dashiell (1993). Knowledge-based interpretation of seismic data in the Intelligent Monitoring System, *Bull. Seism. Soc. Am.* **83**, 1507–1526.
- Cichowicz, A. (1993). An automatic S-phase picker, *Bull. Seism. Soc. Am.* **83**, 180–189.
- Earle, P. S. and P. M. Shearer (1994). Characterization of global seismograms using an automatic-picking algorithm, *Bull. Seism. Soc. Am.* **84**, 366–376.
- Engdahl, E. R. and R. H. Gunst (1966). Use of high speed computer data, *Bull. Seism. Soc. Am.* **78**, 1725–1743.
- Jackson, P. (1990). *Introduction to Expert Systems*, Addison-Wesley, Reading, Massachusetts.
- Kennett, B. L. N. (1991). IASPEI 1991 Seismological Tables, Bibliotech, Canberra.

Table 9
Processing sequence for Event D

Input: Phase	1	2	3	4	5	6	7	8
t_i (sec)	277.0	281.5	437.4	440.5	465.5	653.0	798.6	807.4
a_i	0.023	0.042	0.308	0.238	0.483	0.204	0.087	0.131
v_i (Hz)	1.250	0.212	1.042	0.212	0.521	0.119	0.893	0.169
ϕ_i (°)	136.5	160.3	359.3	145.0	318.7	325.8	333.7	325.9
ψ_i (°)	9.2	9.8	9.9	11.8	11.2	19.1	15.5	0.43
c_i	<i>P</i>	<i>P</i>	<i>P</i>	<i>P</i>	<i>P</i>	<i>S</i>	<i>S</i>	<i>S</i>

Processing:

Step 1: choice of key phases
Key *P*: phase 2; key *S*: phase 6

Step 2: test on differential time
371.5 sec—surviving models: *P–S*, *PP–SKS*, *PKP–SKS*

Step 3: *P–S* pair feature match
Surviving models: *PKP–SKS*

Step 4: <i>P–S</i> Pair	Depth	Dist	Azim	Event Pattern Match Identified Phases	Misfit
<i>PKP SKS</i>	0	114	160 or 340	phase 2: <i>PKP</i> ; phase 6: <i>SKS</i>	0.6423
	100	114	160 or 340	phase 2: <i>PKP</i> ; phase 6: <i>SKS</i>	0.5763
	300	114	160 or 340	phase 2: <i>PKP</i> ; phase 6: <i>SKS</i>	0.5763
	600	128	160 or 340	phase 2: <i>PKP</i> ; phase 3: <i>pPKIKP</i> , phase 5: <i>sPKIKP</i> ; phase 6: <i>SKS</i>	0.2058

Kennett, B. L. N., E. R. Engdahl, and R. Buland (1995). Constraints on seismic velocities in the Earth from travel times, *Geophys. J. Int.* **122**, 108–124.

Kulháněk, O. (1990). *Anatomy of Seismograms*, Elsevier, Amsterdam.

Mykkeltveit, S. and H. Bungum (1984). Processing of regional seismic events using data from small aperture arrays, *Bull. Seism. Soc. Am.* **74**, 2313–2333.

Richter, C. F. (1958). *Elementary Seismology*, Chap. 17, W. H. Freeman, New York.

Ringdal, F. and E. S. Husebye (1982). Application of arrays in the detection, location and identification of seismic events, *Bull. Seism. Soc. Am.* **72**, S201–S224.

Saari, J. (1991). Automated phase picker and source location algorithm for local distances using a single three-component seismic station, *Tectonophysics* **189**, 307–315.

Slunga, R. (1980). International Seismological Data Center. An algorithm for associating reported arrivals to a global seismic network into groups of defining seismic events, FOA Report, C20386-T1.

Tong, C. (1995). Characterization of seismic phases—an automatic analyzer for seismograms, *Geophys. J. Int.* **123**, 937–947.

Tong, C. and B. L. N. Kennett (1995). Towards the identification of later seismic phases, *Geophys. J. Int.* **123**, 948–958.

Research School of Earth Sciences
Australian National University
Canberra ACT 0200, Australia
(C.T., B.L.N.K.)

Manuscript received 5 December 1995.